

PROCEEDINGS OF SPIE

SPIDigitalLibrary.org/conference-proceedings-of-spie

Analyzing long-term performance of the Keck-II adaptive optics system

Ramey, Emily, Lu, Jessica, Yin, Ruoyi, Robinson, Steve, Wizinowich, Peter, et al.

Emily Ramey, Jessica R. Lu, Ruoyi Yin, Steve Robinson, Peter Wizinowich, Sam Ragland, Jim Lyke, Siyao Jia, Shoko Sakai, Abhimat Gautam, Tuan Do, Matthew Hosek Jr., Andrea Ghez, Mark Morris, Eric Becklin, Keith Matthews, "Analyzing long-term performance of the Keck-II adaptive optics system," Proc. SPIE 11448, Adaptive Optics Systems VII, 1144859 (14 December 2020); doi: 10.1117/12.2563252

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2020, Online Only

Analyzing long-term performance of the Keck-II adaptive optics system

Emily Ramey^a, Jessica R. Lu^a, Ruoyi Yin^a, Steve Robinson^a, Peter Wizinowich^b, Sam Ragland^b, Jim Lyke^b, Siyao Jia^a, Shoko Sakai^c, Abhimat Gautam^c, Tuan Do^c, Matthew Hosek, Jr.^d, Andrea Ghez^c, Mark Morris^c, Eric Becklin^c, and Keith Matthews^e

^aAstronomy Department, University of California, Berkeley, CA 94720, USA

^bW. M. Keck Observatory, 65-1120 Mamalahoa Highway, Kamuela, HI 96743

^cDepartment of Physics and Astronomy, University of California Los Angeles, Los Angeles, CA 90095-1547, USA

^dInstitute for Astronomy, University of Hawaii, Honolulu, HI 96822, USA

^eAstronomy Department, California Institute of Technology, MC 249-17, Pasadena, CA 91125, USA

ABSTRACT

We present an analysis of the long-term performance of the W. M. Keck Observatory Laser Guide Star Adaptive Optics (LGS-AO) system and explore factors that influence the overall AO performance most strongly. Astronomical surveys can take years or decades to finish, so it is worthwhile to characterize the AO performance on such timescales in order to better understand future results. Keck Observatory has two of the longest-running LGS-AO systems in use today and represents an excellent test-bed for investigating large amounts of AO data. Here, we use LGS-AO observations of the Galactic Center (GC) from 2005 to 2019, all taken with the NIRC2 instrument on the Keck-II telescope, for our analysis. We combine image metrics with AO telemetry files, MASS/DIMM turbulence profiles, seeing information, and weather data in one cohesive dataset to highlight areas of potential performance improvement and train a simple machine learning algorithm to predict the delivered image quality given current atmospheric conditions. The complete dataset will be released to the public as a resource for testing new predictive control and PSF-reconstruction algorithms.

Keywords: Adaptive optics, machine learning, predictive modeling

1. INTRODUCTION

Since its inception, the use of LGS-AO in astronomy has vastly improved scientific image quality, but it has not yet matched the performance of space telescopes, which are limited only by the instrument performance and the diffraction of their primary apertures. A new wave of ground-based telescopes is currently being planned for the near future (6, 14, 17), but further research must be done on the performance of current adaptive optics systems in order to both avoid known pitfalls and improve upon current performance levels. Once put into practice, such research could improve the long-term stability of AO Systems in general, which is crucial for fields or applications that require observations to take place over the span of years or decades. Any long-term astronomical survey making use of AO could benefit greatly from improved stability, with specific applications including microlensing (16), exoplanet imaging (18), observations of low-mass binary stars (13), and Galactic Center research (7). Though the precise results of this analysis will be specific to the Keck-II telescope at the W. M. Keck Observatory, broader trends in our findings could potentially apply to any AO system in use today.

Similar analyses of AO performance have been done on the Gemini Planet Imager (see Tallis et al. 2018, Tallis et al. 2020) specifically looking at the effects of certain variables on the resulting image quality, and in another study (see Savransky et al. 2018), Machine Learning (ML) techniques are used to predict Gemini image quality given environmental and observational factors. Though exact results are specific to each telescope, these papers provide a jumping-off point for the analysis in this research. As in Savransky et al. (20), our strategy is to first find correlations between parameters in our data and the delivered image quality, and then use the most

strongly correlated data to train and test prediction algorithms. We further investigate instrumental issues / potential performance improvements without the use of ML for more easily interpretable results.

The Keck-II telescope is one of the largest ground-based telescopes currently in operation, and it also has one of the longest-running astronomical applications of a Laser Guide Star Adaptive Optics (LGS-AO) system in the world (25, 27), in operation since 2004, which presents a good opportunity to study the performance of an AO system on a long timescale and under a multitude of different weather and seeing conditions. The goals of this project are twofold: the first is to analyze the performance of the Keck-II LGS-AO system to determine potential problems or improvements, and the second is to develop a preliminary predictive tool for the AO corrected image quality given real-time weather and seeing conditions. To accomplish these goals, we use LGS-AO observations of the Galactic Center (GC) taken with Keck II's Near Infrared Camera (NIRC2), along with the corresponding weather, telemetry, and seeing information to characterize current AO performance and to prototype various performance prediction algorithms using weather and seeing data alone.

Though there have been other in-depth studies of AO performance (4, 12, 26), at the time of this paper, and to the authors' knowledge, there has been no tool developed to predict the quality of an LGS-AO observation in real-time for the Keck-II telescope. If successfully implemented, such a tool would be very useful for observation planning, allowing the prioritization of targets with higher predicted image quality on a given night over those which might yield poorer scientific results.

Table 1: Parameters used in analysis of the Keck AO system.

Variable name	Source	Units	Description
strehl	NIRC2 Image	n/a	Ratio of peak intensity of a point source to peak intensity of a diffraction-limited point source of the same magnitude
fwhm	NIRC2 Image	mas	
rms_err	NIRC2 Header	μm	
itime	NIRC2	sec	Integration time per coadd
coadds	NIRC2	count	Number of coadds per image
az	NIRC2	degrees	Telescope azimuth
airmass	NIRC2		Integrated air density along the line of sight
aolbfwhm	NIRC2		AO LBWFS avg fwhm
lsamppwr	NIRC2	watts	Laser power
lgrmswf	NIRC2		LGSAO RMS WF Residual
MASS	MKWC	arcsec	Integrated MASS seeing
MASSPRO	MKWC	$m^{1/3}$	Average C_n^2 profile
MASSPRO_half, _1, _2, _4, _8, _16	MKWC	$m^{1/3}$	C_n^2 profile at $\frac{1}{2}$, 1, 2, 4, 8, and 16 km above the telescope pupil
DIMM	MKWC	arcsec	DIMM seeing
wind_speed	CFHT	m/s	Mean wind speed
wind_direction	CFHT	degrees	The direction from which the prevailing wind is blowing
temperature	CFHT	$^{\circ}C$	Reading from the MKWC thermometer
relative_humidity	CFHT	%	Humidity as a fraction of total moisture capacity of the air
pressure	CFHT	millibar	Mean atmospheric pressure
AOAOAMED	NIRC2	ADN	AO WFC AOA camera median light
TUBETEMP	NIRC2	$^{\circ}C$	Average of 15 temperate sensors placed around the telescope dome
mjd	NIRC2	days	Modified Julian Date of the NIRC2 observation
MASS_mjd	MKWC	days	Modified Julian Date of the closest MASS measurement
MASSPRO_mjd	MKWC	days	Modified Julian Date of the closest MASSPRO measurement
DIMM_mjd	MKWC	days	Modified Julian Date of the closest DIMM measurement
cfht_mjd	CFHT	days	Modified Julian Date of the closest weather measurement

2. OBSERVATIONS

Our data for this project consists of Keck-II images of the GC obtained using the LGS-AO system and the NIRC2 camera (PI: K. Matthews). The observations are part of a long-running survey to measure the orbits of stars around the GC over a period of roughly 14 years (2005-2019) (11). These include science images and associated AO telemetry files, which record the performance of the AO system throughout the science observation. More information about the observational parameters for the NIRC2 images can be found in references 9, 10, 15, 21, 25, 27, 28, 29, 30, and 5.

To this data, we add meteorological readings from the Canada-France-Hawaii Telescope (CFHT's) weather tower and MASS / DIMM seeing information from the Mauna Kea Weather Center (MKWC) (3) from roughly the same times as the primary NIRC2 observations. The NIRC2 science observations are described in §2.1, the AO telemetry is described in §2.2, details of the CFHT weather data are in §2.3, and the MASS/DIMM seeing data are described in §2.4.

The four data sets are compiled together and parameters are extracted to produce time-series as described in §2.5. This includes image quality metrics such as Strehl ratio and FWHM, AO performance metrics such as RMS wavefront error (WFE) and loop gains, weather data such as humidity and temperature, and atmospheric seeing at different altitudes. Each of the parameters used for analysis is described in Table 1, and the different filtering methods used to investigate the data are described in Table 2.

Table 2: Description of data samples used for analysis and prediction.

Sample #	Sample name	Description	Rows
1	Raw sample	Unfiltered data combined from weather, seeing, and NIRC2 observations as described in Section 2.5. For each NIRC2 image, we record the closest secondary observation (weather, seeing, telemetry) available, filtering out any observations taken more than 100 minutes from the primary (tighter filters are used in subsequent samples).	5949
2	Analysis sample	Data filtered for invalid values in weather (CFHT) and NIRC2 fields. CFHT readings more than 5 minutes from the NIRC2 image, negative wind speed values, or NaN values in any weather or image column are removed. Lines with a negative Strehl ratio or a FWHM < 30 or > 150 mas are also considered invalid.	5609
3	Prediction sample	Data filtered for invalid values in weather, seeing, and NIRC2 fields. Rows with seeing data taken more than 5 minutes from the NIRC2 image or NaN values in any row of the feature data are removed (a list of feature variables can be found in Table 3). All other filters present in sample 2 also apply.	2229
4	Telemetry sample	Data filtered for invalid telemetry and NIRC2 fields. Telemetry files must have the same naming conventions and be started within ~100 seconds of the relevant NIRC2 observations to be considered valid.	1196

2.1 NIRC2 data

All observations used in this research are of the central 10'' region of the Galactic center, approximately centered on Sagittarius A*, the supermassive black hole at the center of the Milky Way (7, 11, 30) (ICRS Coordinates: RA 17 45 40.03599 DEC -29 00 28.1699). NIRC2 is a near-infrared camera on the Keck II telescope that was specifically designed for use with adaptive optics. All images are taken in the Kp-band and in the narrow-field mode of NIRC2. With these settings, NIRC2 had a field of view (FOV) of $\sim 10'' \times 10''$ and a plate scale of 9.952 mas/pixel (21) until 2014, after which the AO system and NIRC2 camera were realigned, changing the plate scale to 9.971 mas/pixel (21). The data is processed through the standard NIRC2 reduction pipeline, which corrects for geometric distortion and differential atmospheric refraction. The Strehl ratio and FWHM are calculated with Keck Observatory's LGSAO data reduction and analysis software (1). These values measure the overall image quality and are used as target variables in both our analysis of the full dataset and in our predictive algorithm. The Strehl ratio refers to the ratio of the peak intensity of a point source to the peak intensity of the same point source if it was observed in the diffraction-limited regime (i.e. without any atmospheric distortion), and the FWHM is simply the width of the Point Spread Function (PSF) of a point source at half its maximum intensity.

2.2 AO Telemetry data

AO Telemetry files are written during each NIRC2 science observation and contain output from the real-time control computer (RTC), which records data on the AO system status and on individual sub-apertures in the wavefront sensor (WFS) and STRAP (System for Tip-tilt Removal with Avalanche Photodiodes) systems. Data is stored in IDL save files (.sav) in two data structures: one for the wavefront sensor data and one for the STRAP tip/tilt data at millisecond time intervals, along with time-independent data on the full AO system. Each of these data structures contains measurements from a collection of sub-apertures at each timestamp. Of these measurements, we make use of the residual wavefront readings from each sub-aperture and their offset centroids, which measure the distance from the focal point to the physical center of the sub-aperture at any given timestamp. Certain telemetry data, such as the low-bandwidth FWHM and RMS Wavefront Error, are also recorded in the corresponding NIRC2 headers.

2.3 CFHT data

The meteorological data for this study comes from the weather tower at the Canada-France-Hawaii Telescope (CFHT) and can be found on the Mauna Kea Weather Center (MKWC) website (3), with samples taken every five minutes. We gather metrics such as pressure, temperature, wind speed and direction, and relative humidity, and use these in both our analysis of current system performance and in designing our predictive algorithm.

2.4 Seeing data

Our seeing data also comes from the Mauna Kea Weather Center archives (3). This seeing information includes both MASS (Multi-Aperture Scintillation Sensor) and DIMM (Differential Imaging Motion Monitor) readings, which both measure the atmospheric seeing, or turbulence strength, in different ways. The seeing is recorded as the atmospheric induced image blurring, in arcseconds of angular resolution, at a wavelength of 500 nm and at zenith (24). The MASS instrument makes measurements based on the principle that the spatial scale of intensity fluctuations is related to the height of the atmospheric layer that produces them. It uses four concentric apertures to measure the intensity fluctuations for a star near zenith and fits a 6-layer atmospheric profile, $C_n^2(z)$, to the data by assuming a Kolmogorov turbulence model (24) and the assumption that the scale of the turbulence is small. A zenith correction is then applied to the results. It should be noted that the MASS instrument does not report $C_n^2(z)$ directly, but instead the integrated turbulence in each layer in the free atmosphere (>0.5 km):

$$\text{MASS}_i = \int_{z_{i,\min}}^{z_{i,\max}} C_n^2(z) w(z) dz \quad (1)$$

Where i is the layer measured and $w(z)$ is the weight function, an idealized response function of the MASS instrument. Turbulence measurements for each layer are reported in units of $\text{m}^{-1/3}$, and an aggregate MASS statistic is reported in arcseconds.

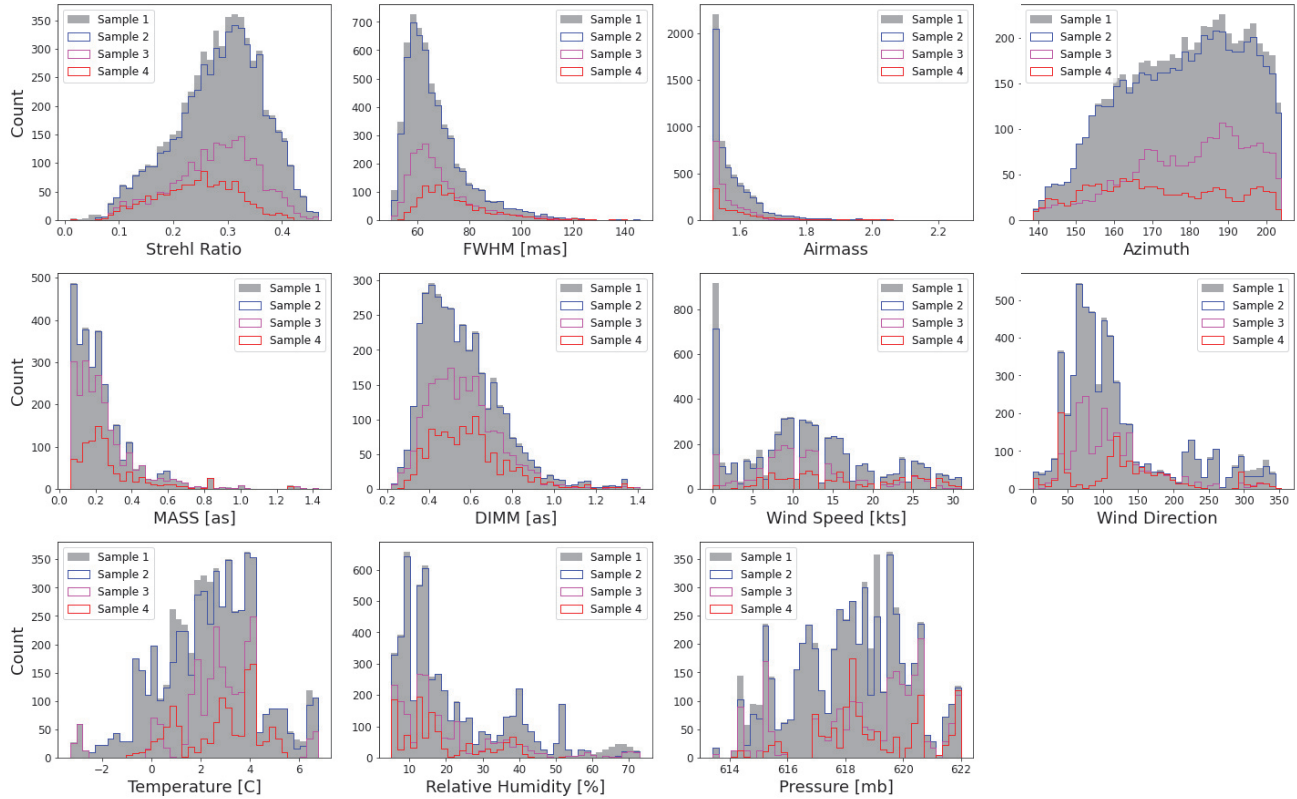


Figure 1: Histograms of feature data before and after filtering. The raw dataset, Sample 1, is shown in gray, while the filtered datasets, Samples 2, 3, and 4, are shown in blue, pink, and red, respectively. The names of the samples and their contents are listed in Table 2.

The DIMM instrument makes a much more straightforward measurement of the seeing, producing one value for the entire integrated atmosphere (up to ≈ 30 km). It measures the overall fluctuation of a single star (not at zenith) and makes a correction for the zenith seeing. Results are reported as the FWHM of a point source for an arbitrarily large telescope at a wavelength of 500 nm (in arcseconds). When the MASS and DIMM instruments are in use, measurements are recorded every few minutes.

2.5 Compiled data

As the weather and seeing data are not taken concurrently with the NIRC2 observations and telemetry files, we must cross-match these files based on their timestamps. The science observations from NIRC2 are used as the primary files, and are linked to the weather and seeing observations with the closest timestamps to their own. The relevant data from each observation type is then extracted to a separate file, as summarized in Table 1, and the data is cleaned in various stages described in Table 2. As the NIRC2 headers themselves include telemetry measurements, we analyze the telemetry files separately from the NIRC2 images until §4.1, in which we compare the NIRC2 header values of telemetry data with calculations from the corresponding telemetry files.

To clean the data, we first remove values which are considered invalid for each sample, as described in Table 2. Some NIRC2 observations are unable to be paired with secondary readings (e.g. observations that predate the MASS and DIMM instruments, had problems with the weather tower, or have no associated telemetry), and this results in null values in weather, seeing, or telemetry fields that must be removed. Observations with negative Strehl ratios and with FWHM below 50 mas or above 150 mas are removed from consideration in all cases as these are nonphysical for NIRC2 images at Kp ($2.2 \mu\text{m}$). Observations with missing seeing or telemetry data are sometimes considered in our analysis of system performance (Section 3.1), but due to the nature of our chosen ML models, any entries with missing values in weather or seeing columns must be taken out before training

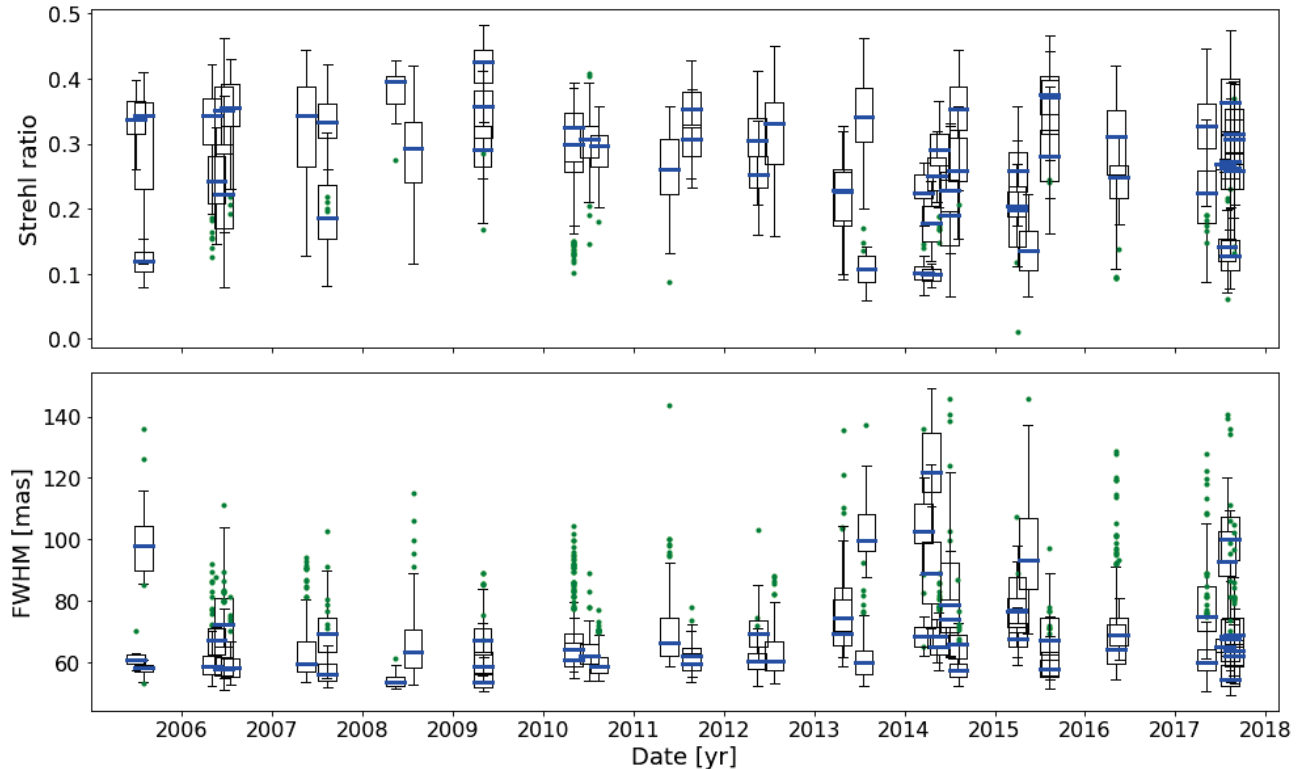


Figure 2: Image quality on each observing night over the full GC survey (sample 2). Boxes indicate the first and third quartiles, with green points showing outliers and medians marked in blue. The variance in image quality increases significantly between 2013 and 2014, with more outliers in later years.

the predictive algorithm. A plot of all Strehl ratios and FWHM values over time, indicating which observations compose which samples, is shown in Figure 1.

3. METHODS

3.1 Performance Analysis

We first perform a visual investigation of the data by plotting relevant seeing and weather parameters against their corresponding image quality metrics, the Strehl ratio and FWHM. We show the Strehl Ratio and FWHM for each observation night over the full GC survey in Figure 2 as a box-and-whisker plot. The plot shows that there have been more outliers in image quality and that the image quality has become more variable in recent years compared to previous observations. After late 2013 or early 2014, there is a higher data spread for each night and higher median FWHM values (lower median Strehl ratios), indicating that the performance of the system has degraded in some way since the survey was started in 2006. This has yet to be confirmed with other instruments on Keck-II or using data from other surveys, however, so the nature of this phenomenon is as yet unknown.

In Figures 3, 4, and 5, we plot seeing statistics, relevant weather statistics, and statistics related to the telescope / observation angle, respectively. The seeing measurements show a visible correlation with their image quality in Figure 3, as expected. However, one odd feature of these plots is that the outdoor and indoor (tube) temperatures (Figure 4) seem to be positively correlated with the Strehl ratio and negatively correlate with the FWHM. This is unexpected, as the LGS-AO system should be able to correct for turbulence at any temperature. We search for other correlations in Figure 6, which shows the Pearson correlation coefficient (Equation 2) of each pair of variables.

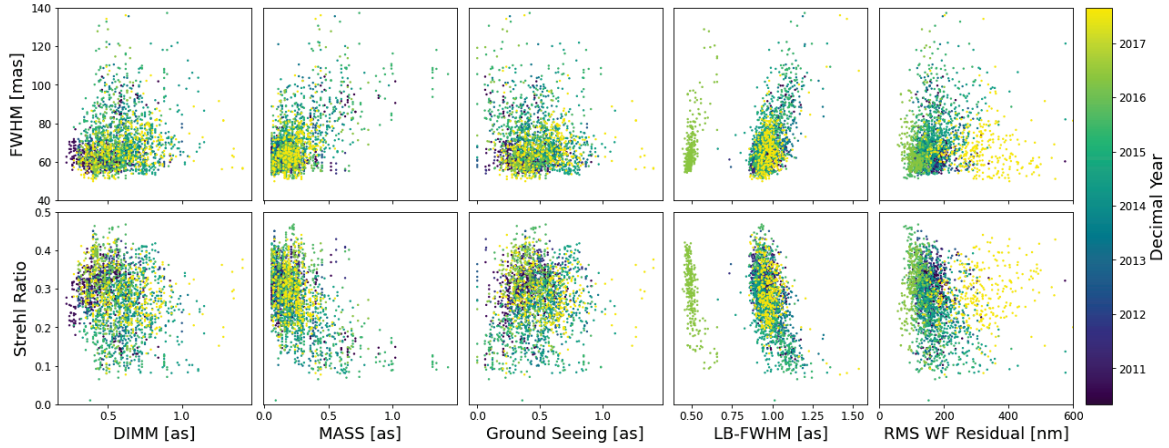


Figure 3: Plot of Sample 3 variables related to seeing and AO correction. As expected, most seeing variables correlate with image quality, but the RMS Wavefront Residual does not, contrary to expectations.

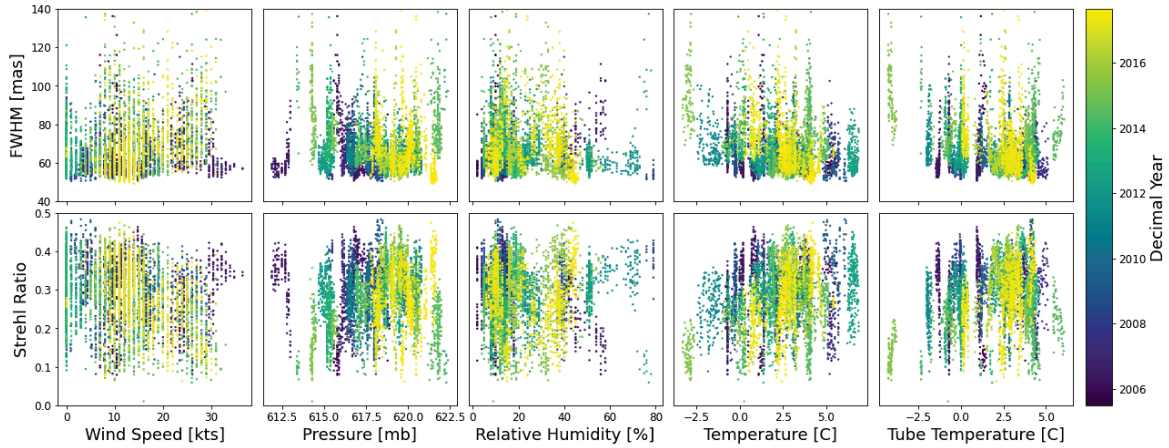


Figure 4: Plot of Sample 2 variables relating to the weather conditions atop Mauna Kea during sample observations. The image quality shows a weak linear dependence on both the outdoor and indoor (tube) temperatures ($r \sim +0.3$ for Strehl, -0.3 for FWHM).

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2)$$

This confirms a correlation between the outdoor temperature and the image quality, which appears to be nearly as strong as the correlation between the image quality and the DIMM seeing. The tubetemp, which is an average of several indoor temperature sensors on the telescope and around the dome, is likewise positively correlated with the Strehl ratio and negatively correlated with the FWHM. A 2018 study of Gemini Planet Imager (GPI) data found a similar dependence of image quality on temperature (22), which will be explored further in Sections 4.1 and 5.

3.2 Predictive Algorithm

The data we use for our analysis of the Keck II AO system (§3.1) is necessarily different from the data we use for our predictive algorithm (§3.2), as the predictive algorithm can only be trained on data that one would have before an observation is made (with the exception of the targets - the Strehl ratio and FWHM).

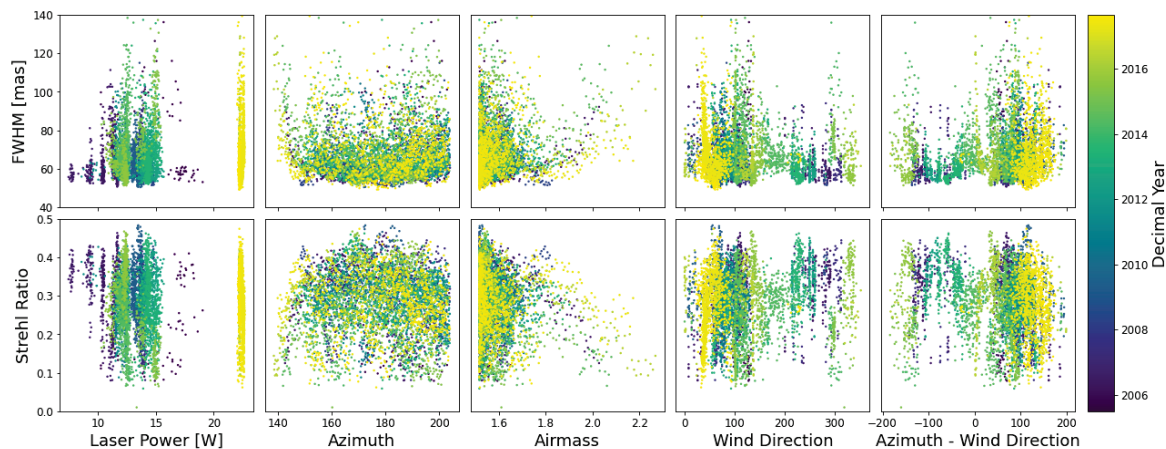


Figure 5: Plot of Sample 2 variables relating to the telescope settings and angle of observation. None of these seem to correlate with image quality.

For example, if we used the RMS Wavefront error to train an algorithm to predict the image quality, we would need to have the RMS Wavefront error every time we wanted to make a prediction, and we would only be able to get that value after an observation, when a prediction would no longer be useful. As the NIRC2 and telemetry data only become available after an observation has taken place, we are limited to data from the CFHT and MASS/DIMM files in building and training our ML model, since these can be obtained in real time.

From the weather and seeing data, we extract only those variables that are expected to have a high impact on overall image quality, shown in Table 3. These variables then become the inputs to our machine learning algorithm, also known as a feature matrix. We continue to use Strehl ratio and FWHM as metrics for the image quality, which become our target variables, or the variables to be predicted by the algorithm. The chosen feature and target variables are listed in Table 3, for reference.

Features	Targets
Wind speed	Strehl ratio
Wind direction	FWHM
Temperature	
Humidity	
Pressure	
Azimuth	
Airmass	
MASS seeing	
DIMM seeing	

Table 3: Feature and target variables for prediction.

3.2.1 Data Transformation

We first standardize the data, which sets the mean of each feature to zero and the standard deviation to one. This is a way to save processing time in models that use a method known as gradient descent in training, and it is an essential step before a Principal Component Analysis (PCA) (8). The PCA algorithm returns a set of orthogonal components, found by taking the eigenvalues of the correlation matrix of the features. Components are listed by their *explained variance*, or how spread out the data is along the given component, and the *explained variance ratio* refers to the explained variance of one component divided by the sum total of explained variances among all the components. While this cannot predict anything from the data on its own, as it does not use the target variables, it has the potential to condense the feature matrix into only a few of its components without losing a significant amount of data variance. Certain machine learning algorithms also require features to be uncorrelated in their underlying assumptions, which can also be achieved through PCA.

We start with a linear PCA - using the feature data as-is, rather than transforming it into a higher-dimensional space first. In order to eliminate a component, we need its explained variance ratio to be below a threshold - 1%, in our case. As shown in Figure 7 (left), none of the components has an explained variance ratio below this threshold, as removing the last component would still eliminate more than 1% of the information contained in our data. This implies that the data is already fairly uncorrelated in the linear regime, but higher-order dependencies are still possible.

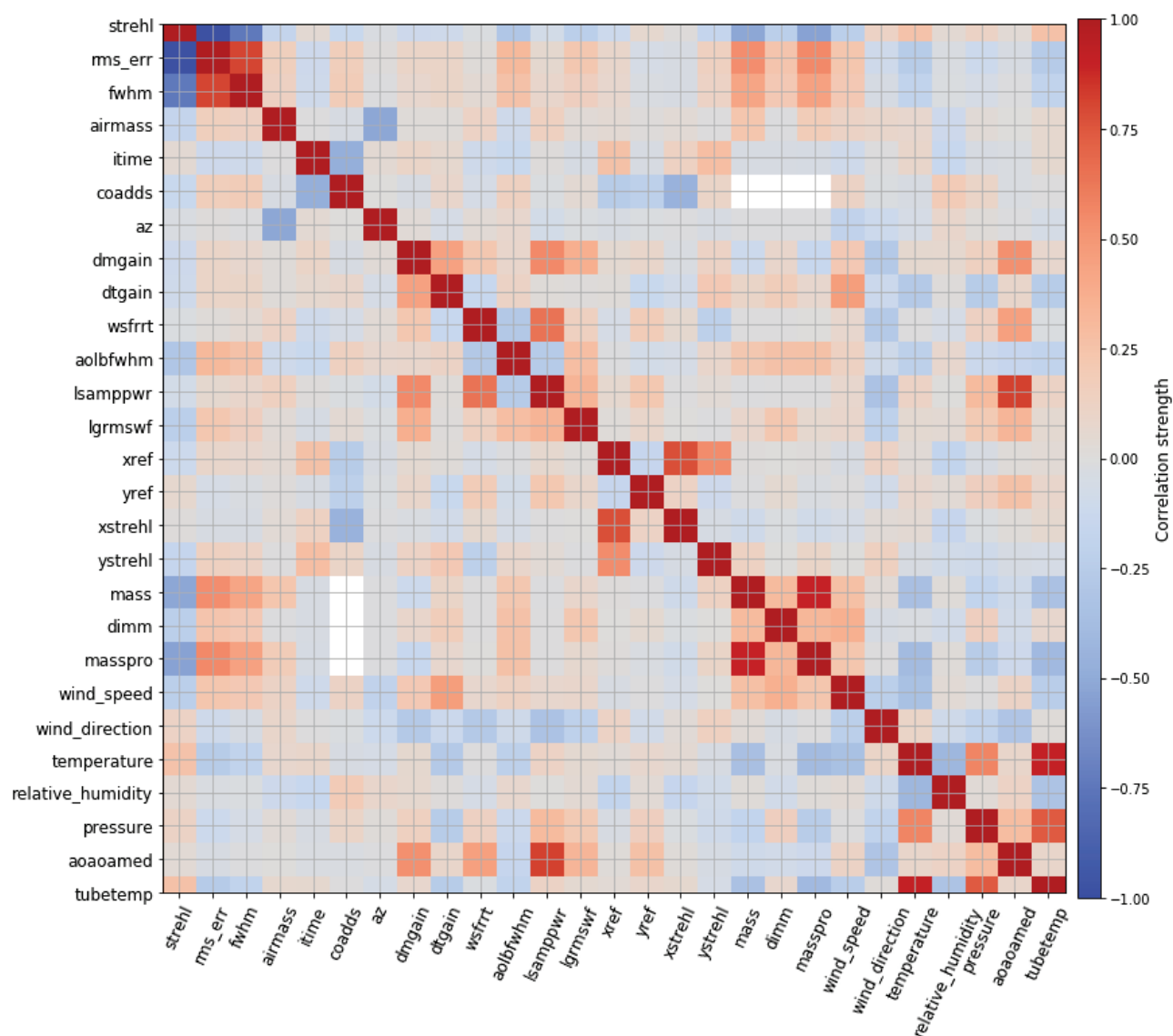


Figure 6: Correlation plot for NIRC2 and weather variables (sample 3). Strong positive correlations are shown in red while strong negative correlations are shown in blue. The strongest correlators with Strehl ratio and FWHM are the temperature, pressure, humidity, and seeing metrics.

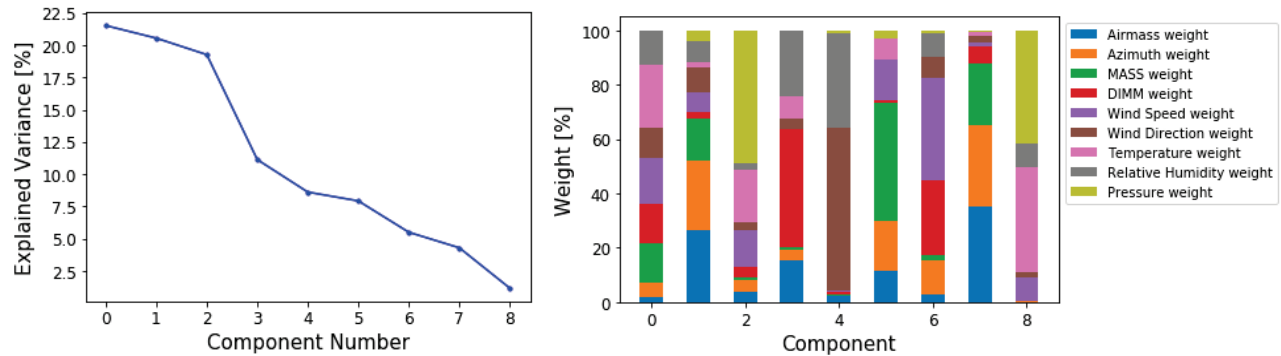


Figure 7: Explained variance ratios (left) and composition of components (right) for a linear PCA on feature variables (Sample 3). No component explains less than 1% or more than 25% of the feature variance, making it difficult to remove any from consideration. The composition of the components reveals that temperature, wind speed / direction, and MASS / DIMM seeing are all large factors in the principal component (component 0), meaning they account for the most significant variation in feature data.

The composition of the components in terms of the original feature data is shown in the right-hand panel of Figure 7. The principal component (component 0) shows that the temperature, wind speed / direction, and MASS / DIMM seeing all play large roles in the feature variance, which confirms the correlations found in Section 3.1. Based on the PCA, we do not eliminate any features from our data in this iteration of the project, although we do use the resulting components in the training of ML models, according to the standard protocol. More advanced methods of feature selection are possible, however, which will be explored in future work.

3.2.2 Models Tested

As we do not know what form the final algorithm should take, we test several machine learning models, largely chosen for their simplicity and/or adaptability to our problem, as well as their value as proofs-of-concept. We have ~ 6000 data points to start with, which reduce to ~ 2000 after the data is cleaned (Table 2). The amount of data is much greater than the number of features, so overfitting is not as much of a problem, and we are less limited in our choice of models as a result. We select three algorithms to start with: Random Forest Regression, Support Vector Regression, and an Artificial Neural Network.

Random Forest Regression is particularly suited to our problem, as it is able to model nonlinear relationships between feature and target data with relative ease. It relies on a simple decision tree as its most basic predictor, and each decision tree is trained on a slightly different subset of the data in order to predict the image quality. The individual results from each decision tree are then averaged to give the final target value. This is a method that is generally very accurate, while also being robust to outliers and overfitting, so it serves as a good test case for our research goals.

Support Vector Regression is an adaptation of a simple Linear Regression, but it makes use of a mathematical technique to maximize the “margins” between the fitted function and the spread of the data. This makes it more robust than a simple linear regression, although both are simple to implement with the `sklearn` package. This model is potentially useful because of a mathematical construct known as the *kernel trick*, which allows it to fit nonlinear functions with a relatively short computation time. One potential drawback to this method, however, is that, in order to use the kernel trick, one must know the approximate shape the solution should take in order to get the best results (e.g. a sine wave, a 4th-order polynomial, etc.).

Artificial Neural Networks (ANN) are a very powerful tool in Machine Learning. They use a series of layers to reduce data, with each layer containing individual “neurons”. Each layer has a series of weights, which are learned in the training process, and a propagation function to transmit data between neurons. Once the desired performance has been reached, the weights are fixed and the network is used only for prediction. This method is intended to solve problems the way a human brain does, and is thus one of the more complicated machine learning models available. Though it is very powerful, it has the potential to overfit the data or to pick up on

Algorithm	Parameter	Values	Step
Random Forest	n_estimators	5-500	10
	max_depth	5-100	5
	max_features	1-#features	1
SVR	degree	1-10	1
	C	0.1-5	0.1
ANN	hidden_layer_sizes	(1)-(10), (1,1)-(20,20)	1

Table 4: Grid search parameters for each algorithm.

correlations that are not present in reality, and great care must be taken in the training and testing process in order to avoid overfitting and get an algorithm that will generalize well to real-world problems.

3.2.3 Model Selection and Validation

In order to compare and evaluate these algorithms, we test the accuracy of each one on data that was not used in the training process. We first split the data into training and test sets and train each model using the training set only. The model is then used to predict the targets for the test set, and the results are compared to the true values to get an estimate of the *test error*, also known as the out-of-sample error or the generalization error, an approximation of how the model will perform in reality. We can take this one step further by using the *cross-validation error*, which involves splitting the data into a number of equal sections (typically 5) and training/testing the algorithm on each resulting train/test split. The results of each test run are then averaged to get the final cross-validation error. This is generally a more robust validation method than simply using the test error, as it is less susceptible to random chance and can provide estimate of the variability of the algorithm's performance. However, all validation methods generally tend to overestimate the true error slightly, as they rely on artificially limiting the size of the training set, which tends to make the algorithm less accurate. This introduces a trade-off in computation time vs. precision of results, as using more splits in a cross-validation is statistically more informative but exponentially more complex.

To compare between models, we start by reporting the results of a simple 80%/20% train/test split, though more complex validation methods are planned for future iterations of the project. Comparing performance within models, however, can be done using the `GridSearchCV` function from the `sklearn` package (19), which internally implements k-fold cross-validation. Each model has a set of *hyperparameters* which determine different things about the fit depending on the model type. For example, a Random Forest algorithm includes such hyperparameters as the number of decision trees used, the maximum depth of each decision tree, and the maximum number of features that can be considered by any individual tree. Choosing the correct hyperparameters is crucial to the machine learning process, and using a grid search method, such as the one mentioned above, is used to test a grid of possible hyperparameter values to find the ones that work best for a given algorithm. A list of the hyperparameter values searched for each algorithm can be found in Table 4. The Grid Search algorithm returns the algorithm with the lowest cross-validation error after it has searched all parameters.

4. RESULTS

4.1 Trends in Weather and Seeing

With Figure 6, we confirm our findings from Figures 3-5 that the Strehl ratio and FWHM correlate with the MASS and DIMM seeing, and we can see that the other weather parameters are very slightly correlated with the image quality as well, even in the linear regime, although we expect the higher-order correlations to be stronger overall. We again see a visible correlation between the outdoor temperature and image quality (Figure 4), which is unexpected given that the telescope and AO system should, in theory, be able to perform well at all temperatures. We see a similar correlation in the tube temperature readings, which reflect an average of several temperature sensors from around the telescope dome. Taking the Pearson correlation (Equation 2) of both temperature measurements with the image quality reveals that both have a correlation coefficient of $r \sim +0.3$ with Strehl ratio and $r \sim -0.3$ with FWHM. This indicates that the system performance worsens as the outdoor temperature becomes colder. Results from the Gemini Planet Imager (4, 22, 23) have revealed

Table 5: Mean Absolute Errors (top of each row) and Mean Absolute Percentage Errors (bottom of each row) for each ML Algorithm tested.

Algorithm	Best Training Error		Best Testing Error	
	Strehl	FWHM	Strehl	FWHM
Random Forest	0.01 4.55%	1.96 mas 2.49%	0.03 17.48%	5.42 mas 6.69%
SVR	0.04 19.27%	6.8 mas 8.48%	0.04 25.29%	7.33 mas 8.71%
ANN	0.04 17.7%	6.38 mas 8.22%	0.04 27.83%	7.15 mas 8.89%

that temperature can be a major factor in the delivered image quality, and that, in that case, the dependence is linked to a temperature difference between the primary mirror and the outside air. Temperature differences anywhere in the telescope's beam path can lead to turbulence, which can be difficult to correct with a standard AO system, as it is calibrated for an atmospheric turbulence profile and not a man-made one, and it is possible that this is exactly what is occurring at Keck-II. However, the additional correlation of the image quality with the dome temperature (tubetemp) indicates that a temperature differential may be occurring somewhere else in the system, instead of in the dome or across the primary mirror, as one feature of the GPI study was a large difference between the dome temperature and the outdoor temperature, which does not seem to be the case with Keck-II.

Further investigation into all potential sources of turbulence would be needed to draw any firm conclusions, but the direction of the temperature correlation (positive with image quality) may also offer some clues as to what is occurring. Since the correlation is consistently positive with image quality, it is likely that the issue is coming from a source that is routinely warmer than the outdoors. At Keck-II, the LGS-AO hardware is housed in an enclosure next to a heated room, with a hatch in the wall leading to the outdoors. Theoretically, heat could be leaking into the enclosure, which would then be exposed to the outdoor air through the AO hatch, allowing cold and warm air to mix and cause turbulence in the enclosure itself. Again, however, more research is needed in order to confirm this theory, which will occur in future iterations of this project.

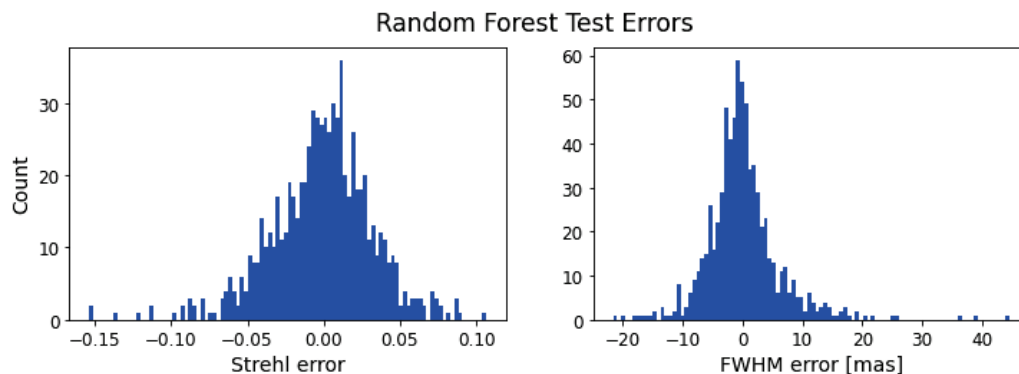


Figure 8: Histogram of Random Forest Strehl and FWHM error values. Errors are distributed roughly evenly about zero, with heavier tails in one direction.

4.2 Predictive Algorithm

After performing a grid search over each algorithm's hyperparameters as described in Table 4, we list the results of the best-performing functions in each category in Table 5. We find that the algorithm with the best overall performance is the Random Forest Regressor. This is logical for our naive approach, as it is the most adaptable of the algorithms tested given the relatively simple feature set and model-selection method. The distribution of errors for both Strehl ratio and FWHM are shown in Figure 8, and Figure 9 contains a plot of the Strehl and FWHM predicted values and percent errors against the true values. We see a sharp increase in Strehl

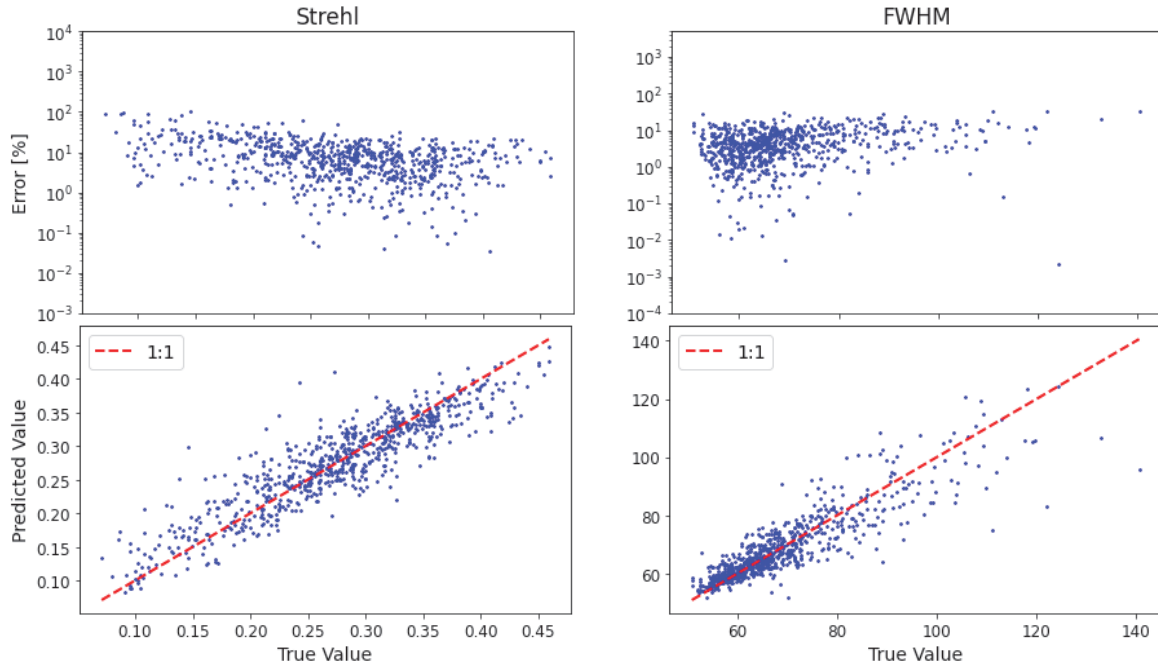


Figure 9: Random Forest Strehl and FWHM predicted values (bottom) and percentage errors (top) plotted against the true values. The predicted Strehl ratio and FWHM are less accurate at smaller values when accuracy is considered as a percentage of the true value.

percentage error as the Strehl ratio itself gets lower, as seen in Figure 9. This could be due to the fact that Strehl calculations themselves become less accurate at lower Strehl (2), which could interfere with the accuracy of our predictive algorithm in that range. These preliminary results are encouraging, however, in that predictions of image quality are accurate to more than 90% on the test FWHM data (80% for Strehl), indicating that a more complex processing scheme, or one that includes more features, has the potential to improve upon this performance.

5. DISCUSSION AND FUTURE WORK

From our analysis, we have found several areas for further investigation and potential improvement of the Keck-II LGS-AO system. The median image quality of the NIRC2 camera (Figure 2) has gotten worse over time, with a significant change in quality occurring just before 2014. Although the causes of this are as yet unknown, the fact that the image quality seems to decline sharply rather than gradually indicates that instrumentation problems are at least partially to blame. Keck Observatory keeps detailed logs of changes to AO instrumentation, so there is a good chance that these will bring us closer to finding the culprit once they are reviewed in light of these findings. It is unclear whether these problems stem from the NIRC2 instrument itself, the AO system overall, or the Keck-II telescope, but a similar analysis of data using other Keck-II instruments or other survey data could clarify this point.

In addition, we have found that the image quality tends to degrade at lower temperatures. Studies of the Gemini Planet Imager, or GPI (see 20, 22, 23), have found that one of the major factors impacting system performance is the temperature difference between the primary mirror and the atmosphere, as it causes turbulence within the telescope dome itself. However, in this case, we see the same correlation with image quality for both the external temperature, measured by the CFHT weather station, and the `tubetemp` variable, which is measured from 15 sensors inside the telescope dome, whereas in the GPI study, it is the temperature difference between the dome and the outdoors that is found to have the largest impact on AO performance. Though the dome or mirror seeing could still be a factor in this case, the direction of the correlations between temperature variables indicates that these are likely not the only factors affecting the image quality. One possible source of the temperature

change could be occurring in the AO enclosure, as it is partially open to the outdoors. In any case, the Keck telescopes could benefit from a more in-depth study of temperature conditions to determine the exact causes of turbulence and the effects on the system.

With the data collected, we were able to train a Random Forest ML model to predict the Strehl ratio to within 0.03 and the FWHM to within 5.4 mas, which is a good starting point for future iterations of this project. Our algorithm performs worse at lower Strehl ratios and higher FWHM values, however, so this is something to address before creating a formal observing tool with this strategy. Other potential improvements include gathering more data as inputs, such as turbulence parameters (r_0 , τ_0 , θ_0), guide star brightness, etc., or training the algorithm on a larger data set, including observations of other targets and other regions of the sky. In Bailey et al. (4), MASS and DIMM seeing are shown to have significant predictive power on the image quality, as is the case in this study, but the variable with the largest predictive power is found to be the Fried parameter r_0 , which can be estimated from the seeing. Future work could significantly benefit from the inclusion of this parameter in the training of our predictive algorithm, as it may improve the final accuracy. A higher level of accuracy for this algorithm, and the eventual development of an observing tool, would mean higher-quality imaging data from Keck-II in the future, as it would enable better target selection during each observing run, particularly for larger surveys which have a greater number of targets to choose from on each night.

References

- [1] Keck LGSAO Software. <https://www2.keck.hawaii.edu/optics/lgsao/software/>, . URL <https://www2.keck.hawaii.edu/optics/lgsao/software/>.
- [2] Keck Strehl Meter. https://www2.keck.hawaii.edu/optics/auchar/Strehl_meter2.htm, . URL https://www2.keck.hawaii.edu/optics/auchar/Strehl_meter2.htm.
- [3] Mauna Kea Weather Center - Home. <http://mkwc.ifa.hawaii.edu/>. URL <http://mkwc.ifa.hawaii.edu/>.
- [4] Vanessa P. Bailey, Lisa A. Poyneer, Bruce A. Macintosh, Dmitry Savransky, Jason J. Wang, Robert J. De Rosa, Katherine B. Follette, S. Mark Ammons, Thomas Hayward, Patrick Ingraham, and et al. Status and performance of the Gemini Planet Imager adaptive optics system. *arXiv:1609.08689 [astro-ph]*, page 99090V, Jul 2016. doi: 10.1117/12.2233172. arXiv: 1609.08689.
- [5] A. Boehle, A. M. Ghez, R. Schödel, L. Meyer, S. Yelda, S. Albers, G. D. Martinez, E. E. Becklin, T. Do, J. R. Lu, and et al. AN IMPROVED DISTANCE AND MASS ESTIMATE FOR SGR A* FROM A MULTISTAR ORBIT ANALYSIS. *The Astrophysical Journal*, 830(1):17, Oct 2016. ISSN 0004-637X. doi: 10.3847/0004-637X/830/1/17.
- [6] Antonin H. Bouchez, D. Scott Acton, Roberto Biasi, Rodolphe Conan, Brady Espeland, Simone Esposito, Josema Filgueira, Daniele Gallieni, Brian A. McLeod, Enrico Pinna, Fernando Santoro, Gelys Tranco, and Marcos A van Dam. The Giant Magellan Telescope adaptive optics program. In Enrico Marchetti, Laird M. Close, and Jean-Pierre Véran, editors, *Adaptive Optics Systems IV*, volume 9148, pages 300 – 318. International Society for Optics and Photonics, SPIE, 2014. doi: 10.1117/12.2057613. URL <https://doi.org/10.1117/12.2057613>.
- [7] Tuan Do, Aurelien Hees, Andrea Ghez, Gregory D. Martinez, Devin S. Chu, Siyao Jia, Shoko Sakai, Jessica R. Lu, Abhimat K. Gautam, Kelly Kosmo O’Neil, and et al. Relativistic redshift of the star S0-2 orbiting the Galactic center supermassive black hole. *Science*, 365(6454):664–668, Aug 2019. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aav8137. arXiv: 1907.10731.
- [8] Karl Pearson F.R.S. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720. URL <https://doi.org/10.1080/14786440109462720>.
- [9] A. M. Ghez, S. D. Hornstein, J. R. Lu, A. Bouchez, D. Le Mignant, M. A. van Dam, P. Wizinowich, K. Matthews, M. Morris, E. E. Becklin, and et al. The First Laser Guide Star Adaptive Optics Observations of the Galactic Center: Sgr A*’s Infrared Color and the Extended Red Emission in its Vicinity. *The Astrophysical Journal*, 635(2):1087, Dec 2005. ISSN 0004-637X. doi: 10.1086/497576.
- [10] A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. Morris, S. Yelda, E. E. Becklin, and et al. Measuring Distance and Properties of the Milky Way’s Central Supermassive Black Hole with Stellar Orbits. *The Astrophysical Journal*, 689(2):1044–1062, Dec 2008. ISSN 0004-637X, 1538-4357. doi: 10.1086/592738. arXiv: 0808.2870.
- [11] Siyao Jia, Jessica R. Lu, S. Sakai, A. K. Gautam, T. Do, M. W. Hosek, M. Service, A. M. Ghez, E. Gallego-Cano, R. Schödel, and et al. The Galactic Center: Improved Relative Astrometry for Velocities, Accelerations, and Orbits near the Supermassive Black Hole. *The Astrophysical Journal*, 873(1):9, Feb 2019. ISSN 1538-4357. doi: 10.3847/1538-4357/ab01de.
- [12] Douglas J. Laidlaw, James Osborn, Timothy J. Morris, Alastair G. Basden, Olivier Beltramo-Martin, Timothy Butterley, Eric Gendron, Andrew P. Reeves, Gérard Rousset, Matthew J. Townson, and et al. Optimizing the accuracy and efficiency of optical turbulence profiling using adaptive optics telemetry for extremely large telescopes. *Monthly Notices of the Royal Astronomical Society*, 483(4):4341–4353, Mar 2019. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty3285. arXiv: 1901.05860.

- [13] Michael C. Liu, Trent J. Dupuy, and Michael J. Ireland. Keck Laser Guide Star Adaptive Optics Monitoring of 2MASS J1534-2952AB: First Dynamical Mass Determination of a Binary T Dwarf. *The Astrophysical Journal*, 689(1):436–460, Dec 2008. ISSN 0004-637X, 1538-4357. doi: 10.1086/591837. arXiv: 0807.0238.
- [14] M. Le Louarn, N. Hubin, M. Sarazin, and A. Tokovinin. New challenges for Adaptive Optics: Extremely Large Telescopes. *Monthly Notices of the Royal Astronomical Society*, 317(3):535–544, Sep 2000. ISSN 0035-8711, 1365-2966. doi: 10.1046/j.1365-8711.2000.03607.x. arXiv: astro-ph/0004065.
- [15] J. R. Lu, A. M. Ghez, S. D. Hornstein, M. R. Morris, E. E. Becklin, and K. Matthews. A DISK OF YOUNG STARS AT THE GALACTIC CENTER AS DETERMINED BY INDIVIDUAL STELLAR ORBITS. *The Astrophysical Journal*, 690(2):1463–1487, Dec 2008. ISSN 0004-637X. doi: 10.1088/0004-637X/690/2/1463.
- [16] J. R. Lu, E. Sinukoff, E. O. Ofek, A. Udalski, and S. Kozlowski. A Search for Stellar-Mass Black Holes via Astrometric Microlensing. *The Astrophysical Journal*, 830(1):41, Oct 2016. ISSN 1538-4357. doi: 10.3847/0004-637X/830/1/41. arXiv: 1607.08284.
- [17] Bruce Macintosh, Mitchell Troy, Rene Doyon, James Graham, Kevin Baker, Brian Bauman, Christian Marois, David Palmer, Donald Phillion, Lisa Poyneer, Ian Crossfield, Philip Dumont, B. Marty Levine, Michael Shao, Gene Serabyn, Chris Shelton, Gautum Vasisht, James K. Wallace, Jean-Francois Lavigne, Philippe Valee, Neil Rowlands, Ken Tam, and Daniel Hackett. Extreme adaptive optics for the Thirty Meter Telescope. In Brent L. Ellerbroek and Domenico Bonaccini Calia, editors, *Advances in Adaptive Optics II*, volume 6272, pages 201 – 215. International Society for Optics and Photonics, SPIE, 2006. doi: 10.1117/12.672032. URL <https://doi.org/10.1117/12.672032>.
- [18] Eric L. Nielsen, Robert J. De Rosa, Bruce Macintosh, Jason J. Wang, Jean-Baptiste Ruffio, Eugene Chiang, Mark S. Marley, Didier Saumon, Dmitry Savransky, S. Mark Ammons, and et al. The Gemini Planet Imager Exoplanet Survey: Giant Planet and Brown Dwarf Demographics from 10 to 100 au. *The Astronomical Journal*, 158:13, Jul 2019. doi: 10.3847/1538-3881/ab16e9.
- [19] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [20] Dmitry Savransky, Jacob Shapiro, Vanessa Bailey, Robert De Rosa, Jason Wang, Jean-Baptiste Ruffio, Eric Nielsen, Melisa Tallis, and Marshall Perrin. Mining the GPIES database. 0703:107030H, Jul 2018. doi: 10.1117/12.2312177.
- [21] M. Service, J. R. Lu, R. Campbell, B. N. Sitarski, A. M. Ghez, and J. Anderson. A New Distortion Solution for NIRC2 on the Keck II Telescope. *Publications of the Astronomical Society of the Pacific*, 128:095004, Sep 2016. ISSN 0004-6280. doi: 10.1088/1538-3873/128/967/095004.
- [22] Melisa Tallis, Vanessa P. Bailey, Bruce Macintosh, Jeffrey K. Chilcote, Lisa A. Poyneer, Jean-Baptiste Ruffio, Thomas L. Hayward, and Dmitry Savransky. Air, Telescope, and Instrument Temperature Effects on the Gemini Planet Imager’s Image Quality. *arXiv:1807.07157 [astro-ph]*, Jul 2018. URL <http://arxiv.org/abs/1807.07157>. arXiv: 1807.07157.
- [23] Melisa Tallis, Vanessa P. Bailey, Bruce Macintosh, Lisa A. Poyneer, Jean-Baptiste Ruffio, Thomas L. Hayward, Fredrik T. Rantakyö, Jeffrey K. Chilcote, and Dmitry Savransky. Effects of Mirror Seeing on High-Contrast Adaptive Optics Instruments. *Journal of Astronomical Telescopes, Instruments, and Systems*, 6(01):1, Jan 2020. ISSN 2329-4124. doi: 10.1117/1.JATIS.6.1.015002. arXiv: 2002.04649.
- [24] A. Tokovinin and V. Kornilov. Accurate seeing measurements with MASS and DIMM. *Monthly Notices of the Royal Astronomical Society*, 381:1179–1189, Nov 2007. doi: 10.1111/j.1365-2966.2007.12307.x.

- [25] Marcos A. van Dam, Antonin H. Bouchez, David Le Mignant, Erik M. Johansson, Peter L. Wizinowich, Randy D. Campbell, Jason C. Y. Chin, Scott K. Hartman, Robert E. Lafon, Paul J. Stomski, Jr., and et al. The W. M. Keck Observatory Laser Guide Star Adaptive Optics System: Performance Characterization. *Publications of the Astronomical Society of the Pacific*, 118(840):310–318, Feb 2006. ISSN 0004-6280, 1538-3873. doi: 10.1086/499498.
- [26] Maaïke van Kooten, Niek Doelman, and Matthew Kenworthy. Impact of time-variant turbulence behavior on prediction for adaptive optics systems. *Journal of the Optical Society of America A*, 36(5):731, May 2019. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.36.000731. arXiv: 1903.06541.
- [27] Peter L. Wizinowich, Jason Chin, Erik Johansson, Stephan Kellner, Robert Lafon, David Le Mignant, Chris Neyman, Paul Stomski, Doug Summers, Roger Sumner, and et al. Adaptive optics developments at Keck Observatory. In *Advances in Adaptive Optics II*, volume 6272, page 627209. International Society for Optics and Photonics, Jun 2006. doi: 10.1117/12.672104. URL <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6272/627209/Adaptive-optics-developments-at-Keck-Observatory/10.1117/12.672104.short>.
- [28] S. Yelda, A. M. Ghez, J. R. Lu, T. Do, L. Meyer, and M. R. Morris. Adaptive Optics Observations of the Galactic Center Young Stars. *arXiv:1208.3466 [astro-ph]*, pages 84470A–84470A–7, Sep 2012. doi: 10.1117/12.927217. arXiv: 1208.3466.
- [29] S. Yelda, A. M. Ghez, J. R. Lu, T. Do, L. Meyer, M. R. Morris, and K. Matthews. PROPERTIES OF THE REMNANT CLOCKWISE DISK OF YOUNG STARS IN THE GALACTIC CENTER. *The Astrophysical Journal*, 783(2):131, Feb 2014. ISSN 0004-637X. doi: 10.1088/0004-637X/783/2/131.
- [30] Sylvana Yelda, Jessica R. Lu, Andrea M. Ghez, Will Clarkson, Jay Anderson, Tuan Do, and Keith Matthews. Improving Galactic Center Astrometry by Reducing the Effects of Geometric Distortion. *The Astrophysical Journal*, 725(1):331–352, Dec 2010. ISSN 0004-637X, 1538-4357. doi: 10.1088/0004-637X/725/1/331. arXiv: 1010.0064.